



## Space-time trajectories of wind power generation: Parameterized precision matrices under a Gaussian copula approach

Tastu, Julija ; Pinson, Pierre; Madsen, Henrik

*Published in:*

Modeling and Stochastic Learning for Forecasting in High Dimensions

*Link to article, DOI:*

[10.1007/978-3-319-18732-7\\_14](https://doi.org/10.1007/978-3-319-18732-7_14)

*Publication date:*

2015

*Document Version*

Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*

Tastu, J., Pinson, P., & Madsen, H. (2015). Space-time trajectories of wind power generation: Parameterized precision matrices under a Gaussian copula approach. In A. Antoniadis, J-M. Poggi, & X. Brossat (Eds.), *Modeling and Stochastic Learning for Forecasting in High Dimensions* (pp. 267-296). Springer. Lecture Notes in Statistics No. 217 [https://doi.org/10.1007/978-3-319-18732-7\\_14](https://doi.org/10.1007/978-3-319-18732-7_14)

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Space-time trajectories of wind power generation: Parametrized precision matrices under a Gaussian copula approach

Julija Tastu, Pierre Pinson and Henrik Madsen

**Abstract** Emphasis is placed on generating space-time trajectories of wind power generation, consisting of paths sampled from high-dimensional joint predictive densities, describing wind power generation at a number of contiguous locations and successive lead times. A modelling approach taking advantage of the sparsity of precision matrices is introduced for the description of the underlying space-time dependence structure. The proposed parametrization of the dependence structure accounts for important process characteristics such as lead-time-dependent conditional precisions and direction-dependent cross-correlations. Estimation is performed in a maximum likelihood framework. Based on a test case application in Denmark, with spatial dependencies over 15 areas and temporal ones for 43 hourly lead times (hence, for a dimension of  $n = 645$ ), it is shown that accounting for space-time effects is crucial for generating skilful trajectories.

## 1 Introduction

The large-scale integration of wind energy into power systems and electricity markets induces operational and management challenges owing to the stochastic nature of the wind itself, with its variability and limited predictability [1]. Forecasting of wind power generation, at various spatial and temporal scales, is generally seen as a

---

Julija Tastu  
Siemens Wind Power, Borupvang 9, 2750 Ballerup, Danmark  
e-mail: [julija.tastu@siemens.com](mailto:julija.tastu@siemens.com)

Pierre Pinson  
Technical University of Denmark, Elektrovej 325-058, 2800 Kgs. Lyngby, Danmark  
e-mail: [ppin@dtu.dk](mailto:ppin@dtu.dk)

Henrik Madsen  
Technical University of Denmark, Matematiktorvet 322-218, 2800 Kgs. Lyngby, Danmark  
e-mail: [hmad@dtu.dk](mailto:hmad@dtu.dk)

crucial input to the decision-making problems involved [23]. An overview of the history of wind power forecasting, though mainly focused on deterministic approaches, as well as an extensive review of the state of the art in that field, are given in [10] and [14], respectively. In parallel, an overview of the current forecasting challenges can be found in [33].

Owing to the complexity of the various decision-making problems at hand, it is preferable that the forecasts provide decision-makers not only with an expected value for future power generation, but also with associated estimates of prediction uncertainty. This calls for probabilistic, rather than point (in the sense of single-valued) forecasting [16]. Example applications of probabilistic forecasting include the design of market offering strategies [6], economic load dispatch and stochastic unit commitment [7, 32], optimal operation of wind-storage systems [9], contingency reserve quantification [4] and assessment of power systems operating costs induced by wind energy [30].

Probabilistic forecasts of wind power generation are commonly generated for each site and lead time of interest, individually. They do not inform about the interdependence structure between potential forecast errors, both in space and in time. Actually, the idea of addressing each site separately can be motivated by the fact that the power curves for the conversion of meteorological variables to power are given by complex non-linear functions of meteorological conditions, number and type of wind turbines, their layout, topographical characteristics, see [22] for instance. Wind power dynamics are generally so site-specific that it is hard to issue high-quality probabilistic forecasts for a large number of sites simultaneously. Similarly, a common practice is to issue power forecasts in the form of marginal predictive densities for each lead time individually, rather than addressing the joint multivariate distribution. The resulting set of marginal predictive densities at  $N$  sites and  $K$  lead times is a suboptimal input for a substantial share of decision-making problems related to power systems operations and electricity markets, e.g., due to power flows on the network or to inter-temporal constraints for conventional power units to be scheduled. A full picture of the space-time characteristics of the stochastic process is there necessary.

Having a set of marginal distributions for a number of random variables, their joint density can be fully characterized using a copula approach. One important feature of copulas is that they can be used to model dependency between random variables independently of their marginal distribution functions. This is important since, as mentioned previously, modelling wind power generation at individual sites while targeting a specific lead time is already a difficult task. It is thus an advantage to decouple the problem of estimating marginal predictive densities from that related to the space-time dependence structure. Copulas have been widely used in many fields for modelling dependencies among sets of random variables, including a number of problems related to wind power. As an example in [5], predictive densities for wind power generation were built by modelling the relation between wind speed and related power output using copulas. In [31], copulas were employed to estimate system net load distribution when accounting for the dependence structure between wind at different locations, at the level of a country, and its relation to the

overall electric consumption. In [20], a copula-based approach was similarly proposed for modelling the spatial dependence in wind speed over the whole European area.

The proposal of issuing trajectories of wind power generation based on predictive marginal densities and a model of their dependence structure was originally described in [36], where the authors focused on a single wind farm, hence considering temporal dependencies only. Based on a Gaussian copula assumption, this temporal dependence was fully specified by an empirical covariance structure to be tracked adaptively and recursively, as in an exponential smoothing framework. Time trajectories of wind power production were then issued by sampling from these multivariate predictive densities. More general parametric and nonparametric approaches were subsequently described in [34], with the main aim of discussing verification of time trajectories. Furthermore, Ref. [26] concentrated on wind power generation at a pair of sites and considered different functional forms of copulas for their dependence, given the lead time. The present study goes along similar lines, even though looking here at the full space-time picture: joint predictive densities of wind power output (and eventually, trajectories) are to be issued based on a set of marginal predictive densities already available for all sites and lead times. The problem then boils down to specifying a model for the dependence structure and to estimating its parameters. Under a Gaussian copula assumption, the modelling approach we propose takes advantage of the sparsity of precision matrices permitting to describe the underlying space-time process. A suitable parametrization of the precision matrix is proposed, hence yielding a more tractable approach even in high dimensions. This proposal goes beyond the conventional assumptions of homogeneous stationary Gaussian Random fields, since the proposed parametrization accounts for the boundary points and considers non-constant conditional variances and direction-dependent conditional correlations.

The paper is structured as following. Sect. 2 introduces the data set used in the study. The methodology is described in Sect. 3. It consists of some preliminaries and definitions, a short introduction to copula modelling and explanation on how precision matrices relate to the Gaussian copula approach. Further, Sect. 4 presents the proposed parametrization of the dependence structure. The estimation method is discussed in Sect. 5, while the empirical results are given in Sect. 6. The paper ends in Sect. 7 with a set of concluding remarks and perspectives regarding future work.

## 2 Data

The case study is for western Denmark, covering the Jutland peninsula and the island of Funen, with a nominal capacity close to 2.5 GW. This corresponds to approximately 70% of the entire wind power capacity installed in Denmark at the time. Even though this nominal capacity regularly evolves due to commissioning and decommissioning of turbines, as well as maintenance operations, it stayed very close to this level over the period considered. Besides the significant share of wind

generation, one of the reasons for concentrating on Denmark relates to its climate and terrain characteristics. The territory is small enough for the incoming weather fronts to affect all of its parts. In addition, the terrain is smooth, therefore passing weather fronts do not meet such big obstacles as mountains when propagating over the country. These aspects make the test case an ideal candidate for understanding space-time effects before moving to more complex cases.

The western Denmark area is divided by the system operator, Energinet.dk, into  $N = 15$  control zones, as depicted in Fig. 1. For all of these areas, power measurements were made available at an hourly resolution. All measurements and related forecasts were normalized by the nominal capacity of the control area they relate to. Consequently, they are expressed in percentage of nominal capacity, generally denoted by  $P_n$ . Point forecasts of wind power generation with an hourly resolution and lead times up to  $K = 43$  hours were produced with the Wind Power Prediction Tool (WPPT) [29]. This corresponds to the most important lead times, today, for power systems operations in an electricity market environment. The update frequency of the forecasts is hourly. Marginal predictive densities were generated in a nonparametric framework, for all control zones and hourly lead times up to  $K$  hours ahead, based on the adaptive resampling approach described in [35]. It comprises one of the state-of-the-art approaches to generating nonparametric predictive densities of wind power generation, in a fashion similar to adaptive quantile regression [37]. These predictive densities are fully characterized by a set of quantile forecasts with varying nominal levels  $\alpha$ ,  $\alpha \in \{0.05, 0.1, \dots, 0.95\}$ . Related predictive cumulative distribution functions are obtained by linearly interpolating through these sets of quantile forecasts.

The available data covers a period from 1 January 2006 to 24 October 2007. For the purpose of the modelling and forecasting study, this dataset was divided into two subsets. The first one covers a period of 11 months, i.e., from 1 January 2006 to 30 November 2006 (8016 forecast series), to be used for the data analysis, model building and estimation. The second subset, covering the period between 1 December 2006 and 24 October 2007 (7872 forecast series), is considered as an evaluation set for the out-of-sample evaluation of the space-time trajectories to be generated, and for comparison with the alternative approaches considered.

### 3 Methodology

The objective of the methodology introduced here is to generate joint predictive densities describing wind power generation at a number of contiguous locations and for a number of successive lead times, independently from the approach used to originally generate the individual predictive densities. These marginal densities are linked together through a (Gaussian) copula function, for which a parametrization of the precision matrix permits to capture the underlying space-time covariance structure. Our proposal methodology can hence be seen as a two-stage approach to the modelling of joint predictive densities, by first obtaining relevant marginal

**Fig. 1** Geographical locations of the  $N = 15$  control areas of Energinet.dk, the system operator in Denmark, for a nominal capacity close to 2.5 GW



predictive densities (here, part of the available data described previously), and then estimating the relevant parameters of the chosen copula. Similar approaches are first instance considered in econometrics applications [19].

### 3.1 Preliminaries and definitions

Let us first describe the general setup for this forecasting problem. At every time step  $t$ , one aims at predicting wind power generation for future times  $t + 1, t + 2, \dots, t + K$  at  $N$  contiguous locations. Seeing wind power generation as a stochastic process, there are in total  $n = NK$  random variables of interest, denoted in the following by  $Y_{t,1}, Y_{t,2}, \dots, Y_{t,n}$ , which we aim at jointly describing given the information available up to time  $t$ . For instance here, with 15 zones and 43 lead times, one has  $n = 645$ . The enumeration is such that  $Y_{t,1}, \dots, Y_{t,K}$  represent wind power generation at the first location for the lead times  $1, \dots, K$ , then  $Y_{t,T+1}, \dots, Y_{t,2T}$  represent wind power generation at the second location for lead times  $1, \dots, K$ , and so on. Uppercase letters are used for random variables, while lowercase letters denote the corresponding observations. Bold font is used to emphasize vectors and matrices. For example,  $\mathbf{y}_t = [y_{t,1}, y_{t,2}, \dots, y_{t,n}]^\top$  stands for the realization of  $\mathbf{Y}_t$ . This translates to generally seeing wind power generation as a vector-valued stochastic process, instead of

a spatio-temporal one. This is more for the sake of employing simpler notations, even though in the following sections the space-time structure will be accounted for, by identifying the spatial and temporal neighbourhood of each random variable composing this vector-valued stochastic process.

The aim of the forecaster is to issue a multivariate predictive distribution  $F_t$ , describing the random vector  $\mathbf{Y}_t = [Y_{t,1}, Y_{t,2}, \dots, Y_{t,n}]^\top$ , conditional to the information available up to time  $t$ ,

$$F_t(y_1, y_2, \dots, y_n) = P(Y_{t,1} \leq y_1, Y_{t,2} \leq y_2, \dots, Y_{t,n} \leq y_n) . \quad (1)$$

Proposing a functional form for  $F_t$  directly implies a simultaneous description of both the marginal densities as well as of the space-time interdependence structure. They should account for the non-Gaussian and bounded nature of wind power generation, as well as for the complex dynamics of wind power variability. Unfortunately, there is no obvious distribution function which could address these required aspects altogether. Employing a copula-based approach appears to be an appealing solution since allowing decomposing the problem of estimating  $F_t$  into two parts.

First, marginal predictive densities,  $F_{t,i} = P(Y_{t,i} \leq y_i)$ ,  $i = 1, 2, \dots, n$ , describing wind power generation at each location and for each lead time individually, can be obtained. In contrast to joint multivariate predictive densities, for which very limited literature exist, the case of issuing marginal ones only is increasingly considered, in both parametric and nonparametric framework. Thus, at this point, the forecaster clearly should take advantage of the state-of-the-art methods available for probabilistic wind power forecasting, while concentrating on an appropriate description of the dependence structure.

Subsequently, the marginal predictive densities can be linked together in order to obtain  $F_t$  using a copula function. Mathematically the foundation of copulas is given by Sklar's theorem [42], which states that, for any multivariate cumulative distribution function  $F_t$  with marginals  $F_{t,1}, F_{t,2}, \dots, F_{t,n}$  there exists a copula function  $C$  such that

$$F_t(y_1, y_2, \dots, y_n) = C(F_{t,1}(y_1), F_{t,2}(y_2), \dots, F_{t,n}(y_n)) . \quad (2)$$

In the case where the joint distribution to be modelled involves continuous random variables only, as for wind power generation, the copula  $C$  is unique.

### 3.2 Copulas for wind power data

Several functional forms of copulas have been considered for wind power data. Namely, in Ref. [36] the authors advocate that a Gaussian copula is an adequate choice when generating joint predictive densities, for a single location and a set of successive lead times. In parallel in Ref. [26], different copula functions were compared for the modelling of the dependence between wind power generation at two

sites, for a given lead time. The results showed that a Gumbel copula performed best, even though Gaussian and Frank copulas could also fit the data fairly adequately.

When moving to higher dimensions, the construction of Archimedean copulas (e.g., Gumbel) becomes complex. For instance, a traditional approach for constructing the  $n$ -variate Gumbel copula requires the  $n^{\text{th}}$  order derivative of the inverse of the process generating function. Even considering explicit formulas for those derivatives given in Ref. [21], the complexity remains high compared to a Gaussian copula approach. Moreover, Ref. [11] showed that in higher dimensions Gaussian copulas outperformed their Gumbel's counterparts, certainly also owing to the much larger potential number of parameters of Gaussian copulas compared to the single parameter the Gumbel ones. Note that these works and results should be interpreted with care as they might depend upon site characteristics, as well as upon the type of marginal predictive densities employed as input.

These works hint at the fact that a Gaussian copula could be deemed appropriate for describing spatial and temporal dependencies present in wind power data. However, these works have not considered spatio-temporal dependencies. Consequently, in a first step, a preliminary data examination is carried out in order to verify whether or not employing a Gaussian copula could be consistent with the space-time dependence structure observed.

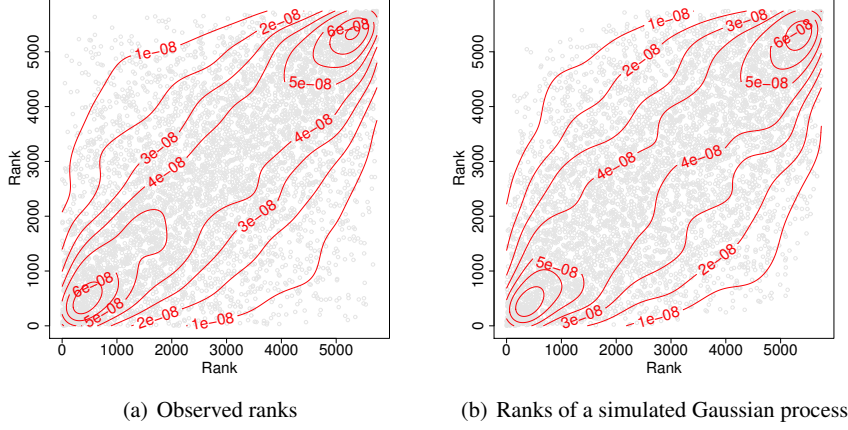
As an example here, consider  $Y_{t,(5 \times 43)+5}$  and  $Y_{t,(4 \times 43)+4}$  representing wind power generation at zone 6 and lead time  $t + 5$ , and wind power generation at zone 5 at lead time  $t + 4$ , respectively. The dependence between the random variables  $Y_{t,(5 \times 43)+5}$  and  $Y_{t,(4 \times 43)+4}$  can be graphically illustrated by focusing on the ranks of the uniform variables  $F_{t,(5 \times 43)+5}(Y_{t,(5 \times 43)+5})$  and  $F_{t,(4 \times 43)+4}(Y_{t,(4 \times 43)+4})$ , for all predictive densities and corresponding realizations available over the first data subset from 1 January 2006 to 30 November 2006.

The scatterplot of the corresponding ranks characterizes the dependence structure between  $Y_{t,(5 \times 43)+5}$  and  $Y_{t,(4 \times 43)+4}$ , while the overlaying contour plot represents the so-called empirical copula [13]. This empirical copula is then compared to what would be the corresponding Gaussian copula, as illustrated in Fig. 2. Both patterns are very similar, thus indicating that the Gaussian copula could be seen suitable for describing the spatio-temporal dependence structure. The results obtained while considering different pairs of variables were all deemed qualitatively similar. Obviously, such a visual comparison does not guarantee that employing a Gaussian copula is the best choice for modelling the dependence structure, while different fit evaluation criteria could be used if really aiming to find an optimal copula (as in [26, 11]). Here it is our choice to focus on Gaussian copulas only, owing to the resulting opportunities given in terms of dependence structure modelling with precision matrices only.

### 3.3 Gaussian Copula

A Gaussian copula is given by





**Fig. 2** Comparison of empirical and Gaussian copula for the dependence between  $Y_{t,(5 \times 43)+5}$  and  $Y_{t,(4 \times 43)+4}$ . The left figure shows a scatterplot with contour overlay for the ranks of  $F_{t,(5 \times 43)+5}(Y_{t,(5 \times 43)+5})$  and  $F_{t,(4 \times 43)+4}(Y_{t,(4 \times 43)+4})$  (empirical copula). In parallel, the right figure depicts a scatterplot with contour overlay for the simulated bivariate Gaussian process having the same rank correlation

$$C(F_{t,1}(y_1), \dots, F_{t,n}(y_n)) = \Phi_{\Sigma}(\Phi^{-1}(F_{t,1}(y_1)), \dots, \Phi^{-1}(F_{t,n}(y_n))) , \quad (3)$$

where  $\Phi^{-1}$  denotes the inverse of the standard Gaussian cumulative distribution function and  $\Phi_{\Sigma}(\cdot)$  is the  $n$ -variate Gaussian distribution function with zero mean, unit marginal variances and a correlation matrix  $\Sigma$ . Based on the assumption such that the marginal predictive densities are probabilistically calibrated, the random variables defined by  $F_{t,i}(Y_{t,i})$ ,  $i = 1, \dots, n$ , are distributed  $U[0, 1]$ . Following an argument similar to that of [28], it is consequently assumed that a joint multivariate predictive density for  $\mathbf{Y}_t$  can be represented by a latent multivariate Gaussian process  $\mathbf{X} = [\Phi^{-1}(F_{t,1}(Y_{t,1})), \dots, \Phi^{-1}(F_{t,n}(Y_{t,n}))]^{\top}$ ,

$$\mathbf{X} \sim \mathcal{N}(0, \Sigma) , \quad (4)$$

with zero mean, unit marginal variances and a correlation matrix  $\Sigma$ . The realizations  $\mathbf{x}_t = [x_{t,1}, \dots, x_{t,n}]^{\top}$  of that process are given by transforming the observations of wind power generation  $y_{t,i}$  through the corresponding predictive cumulative distribution functions and through  $\Phi^{-1}$ ,

$$x_{t,i} = \Phi^{-1}(F_{t,i}(y_{t,i})), \quad i = 1, \dots, n . \quad (5)$$

The base assumption about calibration of marginal predictive densities is core to the methodology subsequently used for modelling the dependence structure. In practice, it might be very difficult to verify whether these densities are calibrated or not,

especially for small samples. Lack of calibration would necessarily translate to  $\mathbf{X}$  not being multivariate Gaussian.

Note that, in this setup, even though the marginal distributions  $F_{t,i}$  as well as the joint distributions  $F_t$  are time-dependent, the underlying dependence structure is fully characterized by the time-invariant correlation matrix  $\Sigma$ , hence not requiring a time index for either  $\mathbf{X}$  or  $\Sigma$ . Such an assumption may not always hold in practice, since the dependence between the sites and different lead times might change under the influence of meteorological conditions, seasonal effects, changes in the terrain roughness, etc. It is out of scope in this study to address those issues, even though we discuss in Sect. 7 several possible extensions permitting to better capture such variations in the spatio-temporal dependence. The most straightforward one would be to use a sliding-window estimation approach, even though it would clearly increase computational costs.

### 3.4 Modelling as a conditional autoregression

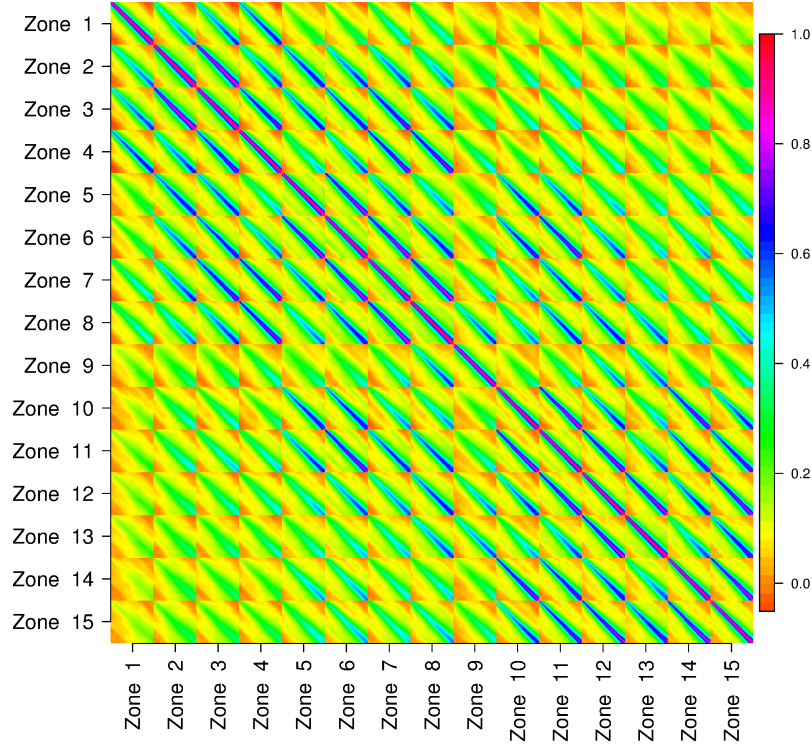
Consider a set of available wind power observations for the vector valued process  $\mathbf{Y}_t$ ,  $t = 1, \dots, T$ . This process is transformed so as to obtain the latent multivariate Gaussian one  $\mathbf{X}$ , and related observations. Emphasis is placed on the correlation structure of  $\mathbf{X}$ . As can be seen from Fig. 3, the sample correlation matrix  $\hat{\Sigma}$  is dense. This directly implies that inference with such a matrix has a computational complexity of  $\mathcal{O}(n^3)$ . In order to make the proposed methodology applicable for problems of high dimension, instead of modelling the covariance matrix directly, we focus on its inverse, the precision matrix, denoted by  $\mathbf{Q}$  [41].

In contrast, the sample precision matrix (see Fig. 4) is very sparse. This suggests considering Gaussian Markov Random Fields (GMRF), allowing us to benefit from computationally efficient algorithms derived for inference with sparse matrices. More specifically, by switching from a dense correlation matrix to its sparse inverse, we reduce the computational complexity from  $\mathcal{O}(n^3)$  to a range from  $\mathcal{O}(n)$  to  $\mathcal{O}(n^{3/2})$ , depending on the process characteristics [41].

While a correlation structure tells of global dependencies between marginal dimensions of the vector-valued process, the precision matrix represents conditional interdependencies. The elements of the precision matrix have the following interpretation. The diagonal elements of  $\mathbf{Q}$  are the conditional precisions of  $X_i$  given  $\mathbf{X}_{-i} = [X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n]^\top$ , while the off-diagonal elements, with a proper scaling, provide information about the conditional correlations between the variables. For a zero mean process such as the one we are dealing with, one has

$$\mathbb{E}(X_i | \mathbf{X}_{-i}) = -\frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij} X_j, \quad (6)$$

$$\text{Var}(X_i | \mathbf{X}_{-i}) = 1/Q_{ii}. \quad (7)$$



**Fig. 3** Sample correlation matrix  $\hat{\Sigma}$  over the first subset of data

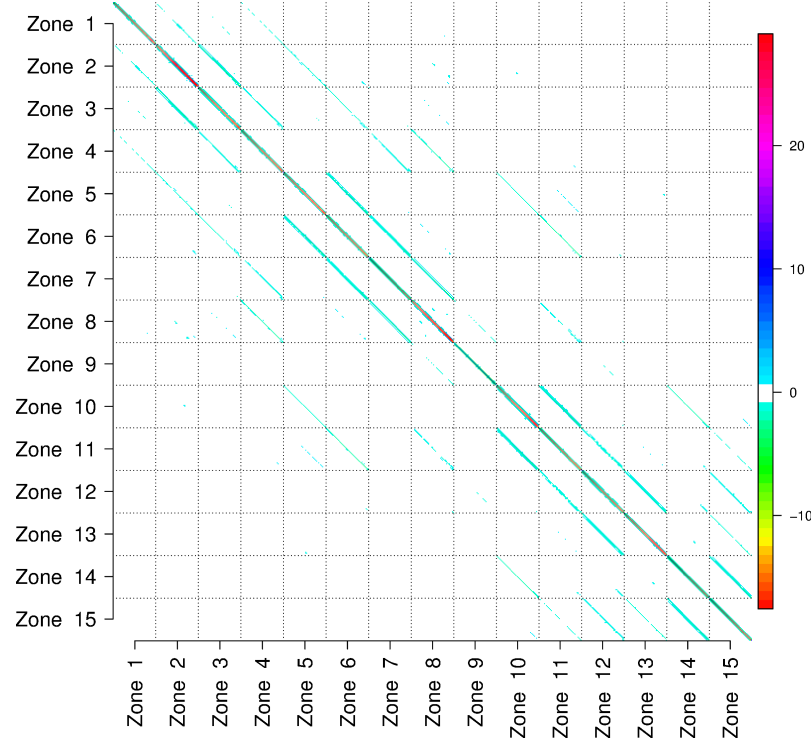
A very important relation is that  $Q_{ij} = 0$  if and only if the elements  $X_i$  and  $X_j$  are independent, given  $\mathbf{X}_{-\{i,j\}}$ . Hence, the non-zero pattern of  $\mathbf{Q}$  determines the neighbourhood of conditional dependence between variables. This relationship can be used to propose a parametrization of the precision matrix. Of course, one still has to keep in mind that  $\mathbf{Q}$  is required to be symmetric positive-definite (SPD).

The relationship given by (6) and (7) is sometimes used for an alternative specification of Gaussian Markov Random Field through the full conditionals. This approach was pioneered by Besag in [3]: the resulting models are also known as conditional autoregressions, abbreviated as CAR. When specifying GMRF through CAR, instead of considering the entries of the precision matrix  $\mathbf{Q}$ ,  $Q_{ij}$ , directly, focus is on modelling terms  $\kappa_i = Q_{ii}$  and  $\beta_{ij} = Q_{ij}/Q_{ii}$ ,  $i, j = 1, \dots, n$ .

From (6) it is seen that the elements  $\beta_{ij}$  are given by the coefficients of the corresponding conditional autoregression models, while  $\kappa_i$  inform on the related variances. This translates to

$$\mathbf{Q} = \kappa \mathbf{B}, \quad (8)$$

where  $\kappa$  denotes a diagonal matrix of dimension  $n \times n$ , the diagonal elements of which are given by  $\kappa_i$ ,  $i = 1, \dots, n$ .  $\mathbf{B}$  is a coefficient matrix gathering a set of co-



**Fig. 4** Sample precision matrix  $\hat{\mathbf{Q}}$  over the first subset of data

efficients  $\beta_{ij}$ ,  $i, j = 1, \dots, n$ , to be seen as a standardized precision matrix. Such a CAR specification is generally easier to interpret and we will use it to propose a parametrization for  $\mathbf{Q}$  in this work.

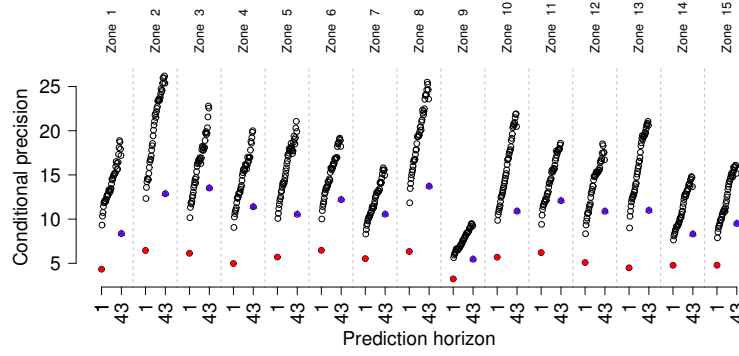
## 4 Parametrization of the precision matrix $\mathbf{Q}$

As the CAR specification of (6) and (7) allows us to decouple the problem of describing  $\mathbf{Q}$  into the matrix of conditional precisions  $\kappa$  and the coefficient matrix  $\mathbf{B}$  is presented, their parametrization are presented one after the other below.

### 4.1 Parametrization of $\kappa$

Conventionally, CAR models are given by stationary GMRF. Stationarity implies rather strong assumptions on both the neighbourhood structure and the elements of

**Q.** Firstly, the structure of the neighbourhood does not allow for special arrangements for the boundary points. Secondly, the full conditionals have constant parameters not depending on  $i$ , i.e., the conditional precisions  $\kappa_i$ ,  $i = 1, \dots, n$ , are assumed constant. However, our data analysis showed that this assumption was too restrictive in the present case. Indeed, having a closer look at the diagonal of the sample precision matrix  $\hat{\mathbf{Q}}$ , depicted in Fig. 5, it is clear that its elements are not constant. Instead, except for the boundary points, they exhibit a trend with conditional precision increasing with the lead time.



**Fig. 5** Diagonal elements of the sample precision matrix,  $\hat{\mathbf{Q}}$ . Boundary points given by the conditional precisions for lead times 1 and 43 hours ahead are marked with red and blue circles, respectively

In addition, the variation patterns for conditional precisions appear similar for all zones. Only the variation pattern for zone 9 looks different. This result is in line with a previous analysis in [43] of the spatial and temporal dynamics of wind power generation for this dataset. On the one hand, that difference might be explained by the smaller area covered by zone 9 in comparison with all the others (and lower nominal capacity), hence leading to more significant local variations, and then lower conditional precisions. On the other hand, that zone also exhibits different characteristics since located off the mainland of Jutland, where offshore wind dynamics can substantially differ from those observed onshore.

If looking at the other zones, the variation patterns observed for the conditional precisions are rather similar. It was not possible to link the differences between zones to the explanatory variables available, either measurements or forecasts. In parallel, even though one might think that the precision pattern could depend on whether a zone is located in the center of the considered territory or on the boundary, such an assumption was not supported by the data. Furthermore, our analysis did not support the alternative assumption such that conditional precisions could depend on the overall level of power variability at that zone. Consequently, our pro-

posed parametrization for conditional precisions is the same for all zones. Further investigation might allow to refine this proposal.

As a result,  $\kappa$  is a block diagonal matrix,

$$\kappa = \begin{pmatrix} \kappa_B & & 0 \\ & \kappa_B & \\ & & \ddots \\ 0 & & & \kappa_B \end{pmatrix}, \quad (9)$$

where its block element,

$$\kappa_B = \begin{pmatrix} \kappa_1 & & 0 \\ & \kappa_2 & \\ & & \ddots \\ 0 & & & \kappa_K \end{pmatrix}, \quad (10)$$

is a diagonal matrix of dimension  $K \times K$ , repeating  $N$  times.

Focusing on a single block element  $\kappa_B$ , and in view of our observations related to Fig. 5, a parametrization for  $\kappa_B$  ought to consider separately the central lead times and the temporal boundaries. These temporal boundaries  $\kappa_1$  and  $\kappa_K$  cannot be avoided since there cannot be lead times of less than 1 hour ahead and more than  $K$  hours ahead. As a result for these lead times, the conditional models in (6) and (7) rely on a smaller set of explanatory variables. This in turn leads to lower precision values.

For lead times between 2 to  $K - 1$  hours ahead, an analysis of Fig. 5 suggests that conditional precisions increase with the lead time, and could be expressed as

$$\kappa_i = \rho^{i-2}, \quad i = 2, \dots, K-1, \quad (11)$$

with  $\rho$  a ratio parameter. This model was found to be more suitable than a simpler linear one, which, visually based on Fig. 5 could also be seen as a relevant candidate.

As a result, this block element  $\kappa_B$  is parametrized as

$$\kappa_B = \frac{1}{\sigma^2} \begin{pmatrix} \kappa_1 & & & 0 \\ & 1 & & \\ & & \rho & \\ & & & \ddots \\ 0 & & & & \rho^{K-2} \\ & & & & & \kappa_K \end{pmatrix} \quad (12)$$

meaning that, overall, the diagonal elements of  $\mathbf{Q}$  can be described with four parameters only: the temporal boundaries  $\kappa_1$  and  $\kappa_K$ , the ratio parameter  $\rho$ , and an overall scaling  $\sigma^2$ .

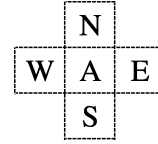
## 4.2 Parametrization of the standardized precision matrix $B$

Our next step is to describe a parametrization for  $B$ . This requires an analysis of the neighbourhood structure of  $Q$  by identifying which elements are non-zero. We first look at the spatial neighbourhood and then at the temporal one. The final parametrization for  $B$  is finally presented.

### 4.2.1 Spatial neighbourhood

Consider a single zone denoted by  $A$ . A careful look at Fig. 4 (while also remembering the spatial distributions of these zones shown in Fig. 1) reveals that information at zone  $A$  only depends on the local information at  $A$  and on the four closest neighbouring zones: northern ( $N$ ), eastern ( $E$ ), southern ( $S$ ) and western ( $W$ ) neighbours of  $A$ . A simplified representation of their spatial neighbourhood effects is given in Fig. 6.

**Fig. 6** Neighbourhood specification of a single zone. The focus zone is marked  $A$ , while  $W$ ,  $N$ ,  $E$  and  $S$  denote its western, northern, eastern and southern neighbours, respectively

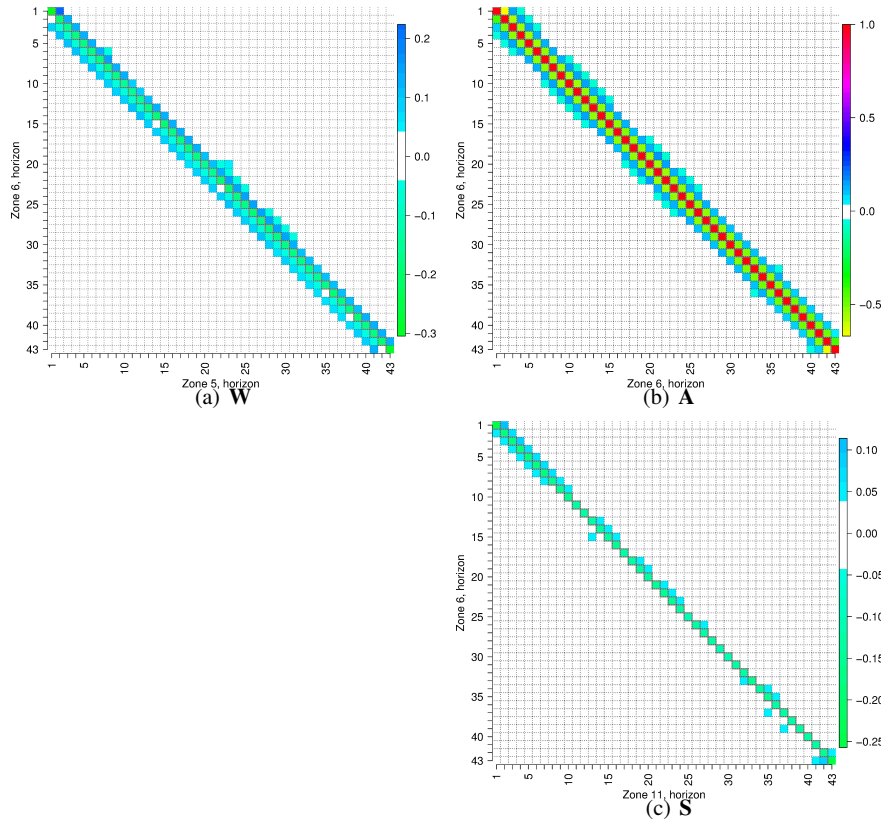


In practice, the spatial neighbours are assigned to each of the zones based on the visual inspection of the spatial distributions of these zones (see Fig. 1), as well as expert knowledge. For instance, for zone 11, the obvious spatial neighbours are  $W=10$ ,  $E=12$  and  $S=15$ . The  $N$  neighbour is more difficult to define, since a unique zone is to be picked, while here both zones 6 and 7 could be seen as appropriate. An analysis of the sample precision matrix allowed to decide on  $N=6$  in view of a more pronounced dependence.

### 4.2.2 Temporal neighbourhood

Fig. 4 shows that information observed at zone  $A$  at time  $t$  is only dependent on a very small amount of elements at zones  $A$ ,  $N$ ,  $E$ ,  $S$ , and  $W$ . Since precision matrices ought to be symmetric, it is sufficient to focus on the dependency between  $A$  and its western and southern neighbours, without direct consideration of the eastern and northern neighbours. Let us zoom into some relevant blocks of the sample coefficient matrix  $\hat{B}$ , obtained based on our first subset of data, when focusing on zone 6.

Based on Fig. 7, it appears that the corresponding conditional correlations of zone A with its northern and the western neighbours differ. Information at zone A observed at time  $t$  is conditionally dependent only on the simultaneous information at zone N. Meanwhile, the conditional correlation with zone W is significant at times  $t - j$ ,  $j = -2, \dots, 2$ . This difference in the dependency patterns can be partly explained by the fact that in Denmark the prevailing winds are westerly. Thus, forecast errors most often propagate from West to East, as discussed in, e.g., [15]. This means that usually zones A and N are influenced by the upcoming weather front simultaneously, while zone W is exposed to it earlier. Of course, one should also keep in mind that in our test case distances between zones A and N are in general larger than those between A and W. That can be another factor influencing the differences.



**Fig. 7** Zooming on blocks of the standardized sample precision matrix  $\hat{\mathbf{B}}$

In general, the results depicted in Fig. 7 show that information corresponding to lead time  $k$  for zone A is dependent on the variables at the neighbouring zones corresponding to lead times  $k - j$ , where  $j = -2, \dots, 2$ . Thus, visually the data suggests a second order (temporal) process. In this work both the second ( $j = -2, \dots, 2$ ) and



the first order ( $j = -1, 0, 1$ ) models have been considered. Since the corresponding difference in the performance of the resulting predictive densities was rather minor, in this study the focus is on the first order model ( $j = 1$ ). Extension to higher order models is rather straight-forward and all the discussed parametrization and estimation procedures apply.

In this work a directional non-stationary CAR model, abbreviated as DCAR, is considered. That is, the conditional correlations are made direction-dependent. In this respect the work is inspired by [24] where the authors consider a directional (in space) CAR model. We refer the reader to that work for a clear description of the modelling approach. The current proposal can be viewed as a generalization of the work presented in [24] since space-time neighbourhoods are considered along with the non-constant precisions.

When considering DCAR models, directional neighbourhoods should be chosen carefully so that each of them forms a (directional) clique. Let us consider two elements from the full random vector  $\mathbf{X}$ ,  $X_i$  and  $X_j$ . Then, given that  $X_i$  is a “west-side” and “one-hour-ago” neighbour of  $X_j$ ,  $X_j$  should be assigned as the “east-side” and “one-hour-ahead” neighbour of  $X_i$ . This is essential for ensuring the symmetry of the precision matrix.

#### 4.2.3 Resulting space-time parametrization for $\mathbf{B}$

Our analysis over the first data subset suggested that information for zone A at lead time  $k$  conditionally depends on information from zones N, E, S, and W with lead times  $k - 1$ ,  $k$ ,  $k + 1$ , as well as on the local situation at zone A for lead times  $k - 1$  and  $k + 1$ . In terms of the CAR specification given in (6) this translates to

$$\begin{aligned} \mathbb{E}(X_{t,(A-1)K+k} | \mathbf{X}_{t,-\{(A-1)K+k\}}) = & - \sum_{j=\{-1,1\}} a_j X_{t,(A-1)K+(k+j)} \\ & - \sum_{j=\{-1,0,1\}} (b_j X_{t,(W-1)K+(k+j)} + b_j^* X_{t,(E-1)K+(k+j)} \\ & + c_j X_{t,(N-1)K+(k+j)} + c_j^* X_{t,(S-1)K+(k+j)}) . \end{aligned} \quad (13)$$

In the above,  $a_j$ ,  $b_j$ ,  $b_j^*$ ,  $c_j$  and  $c_j^*$  denote the coefficients representing conditional dependencies on the local ( $a$ ) information as well as on the information coming from the west ( $b$ ), east ( $b^*$ ), north ( $c$ ) and south ( $c^*$ ), respectively. These coefficients are the building blocks for  $\mathbf{B}$ .

It was found that  $a_j$ ,  $b_j$ ,  $b_j^*$ ,  $c_j$  and  $c_j^*$  do not depend on the considered lead time  $k$ . It can be also seen from Fig. 7 that there is no indication of any increase/decrease of the coefficient values with the lead time. The only values which differ from the general picture are those for the temporal boundaries. This effect is already dealt with when scaling by the corresponding conditional precisions.

In this work it is assumed that the corresponding coefficients are constant for all zones. Further work could be done in order to explain spatial variations in the coefficient values. Some restrictions have to be imposed on the parameters in  $\mathbf{B}$  to

For the specific setup of our Danish test case, the resulting standardized precision matrix is defined as

where  $\mathbf{W}$ ,  $\mathbf{N}$  represent the blocks describing the conditional dependencies between a given zone its Western and Northern neighbours, respectively, while  $\mathbf{A}$  represent the local dependencies within zone  $A$  itself. Note that indices for the zones were added in order to better appraise the structure of this matrix.

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{a_{-1}}{\kappa_1} & & & 0 \\ a_{-1} & 1 & \rho a_{-1} & & \\ & \ddots & \ddots & \ddots & \\ & & a_{-1} & 1 & \rho a_{-1} \\ 0 & & & \frac{\rho^{K-1} a_{-1}}{\kappa_K} & 1 \end{pmatrix}. \quad (15)$$

One can note that in the upper diagonal of  $\mathbf{A}$  instead of writing  $a_1$  as suggested by (13), we directly express  $a_1$  in terms of  $a_{-1}$  in order to ensure symmetry of the resulting  $\mathbf{Q}$ . That is, the upper diagonal of  $\mathbf{A}$  when multiplied by  $[\kappa_1, 1, \rho, \dots, \rho^{K-2}]^\top$  has to be equal to the lower diagonal of  $\mathbf{A}$  multiplied by  $[1, \rho, \dots, \rho^{K-2}, \kappa_K]^\top$ . From this one can directly obtain the dependency between the upper and the lower diagonals of  $\mathbf{A}$ . Since the elements of  $\kappa_B$  increase proportionally by  $\rho$  for lead times from 2 to  $K-1$ , then  $a_1 = \rho a_{-1}$  in the corresponding part. In the similar manner, the scaling for the boundary points is a function of  $\kappa_1$  and  $\kappa_K$ .

$$\mathbf{W} = \begin{pmatrix} b_0 & \frac{b_1}{\kappa_1} & & & 0 \\ b_{-1} & b_0 & b_1 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{-1} & b_0 & b_1 \\ 0 & & & \frac{\rho^{K-1} b_{-1}}{\kappa_K} & b_0 \end{pmatrix}, \quad (16)$$

and

$$\mathbf{N} = \begin{pmatrix} c_0 & \frac{c_1}{\kappa_1} & & & 0 \\ c_{-1} & c_0 & c_1 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{-1} & c_0 & c_1 \\ 0 & & & \frac{\rho^{K-1} c_{-1}}{\kappa_K} & c_0 \end{pmatrix}. \quad (17)$$

For the last two blocks, they are readily obtained with

$$\begin{aligned} \mathbf{E} &= \mathbf{K}^{-1} \mathbf{W}^\top \mathbf{K}, \\ \mathbf{S} &= \mathbf{K}^{-1} \mathbf{N}^\top \mathbf{K}, \end{aligned} \quad (18)$$

to ensure symmetry of  $\mathbf{Q}$ .

Finally, the precision matrix  $\mathbf{Q}$  can be fully modelled by a parameter vector  $\theta$ , as

$$\theta = [\kappa_1, \rho, \kappa_K, \sigma^2, a_{-1}, b_0, b_{-1}, b_1, c_0, c_{-1}, c_1]^\top. \quad (19)$$

It will hence be referred to as  $\mathbf{Q}(\theta)$  in the following.

## 5 Estimation

We explain here how to fit the GMRF defined by  $\mathbf{Q}(\theta)$  to the observations. This task can be divided into two parts. Firstly, the discrepancy measure between observed data and the suggested GMRF is to be chosen. Secondly, one has to insure that the parameter estimates belong to the valid parameter space  $\Theta^+$ , such that the resulting precision matrix is SPD.

### 5.1 The valid parameter space

The precision matrix  $\mathbf{Q}$  was previously described as a function of the parameter vector  $\theta$ . Symmetry of  $\mathbf{Q}(\theta)$  is imposed by construction. Hence, we are left with the issue of the positive definiteness of  $\mathbf{Q}(\theta)$ .

Unfortunately, in general, it is hard to determine the valid parameter space  $\Theta^+$ . Analytical results are available for precision matrices that are Toeplitz [39]. These results can be used when working with homogeneous stationary GMRFs, but this is not the case here. An alternative to consider here is to work with a subset  $\Theta^+$  given by the sufficient condition such that  $\mathbf{Q}(\theta)$  is diagonal dominant.

Diagonal dominance is easier to treat analytically. On the downside, this approach becomes more and more restrictive for an increasing number of parameters. This issue is discussed in more detail in [39]. For instance, for our particular test case we could see that the assumption of diagonal dominance was too restrictive. And, if no such restriction was imposed, the vector of parameters estimated was far from yielding a diagonal dominant precision matrix.

If the full valid parameter space  $\Theta^+$  is unknown and its diagonal dominant subset is deemed too restrictive, it is always possible to use a “brute force” approach (following the terminology of [39]). This entails checking if  $\hat{\theta} \in \Theta^+$  by direct verification of whether the resulting  $\mathbf{Q}(\hat{\theta})$  is SPD or not. This is most easily done by trying to compute the Cholesky factorization which will be successful if and only if  $\mathbf{Q}$  is positive definite. This is the approach we decided to use here.

However, it is worth mentioning some advantages given by the diagonal dominance approach over the “brute force” method. An important one is that if one estimates parameters while requiring the diagonal dominance, then one can be sure that if a new territory is to be included, there is no strict necessity (other than aiming for optimality) to re-estimate the parameters. In other words, one can be sure that the vector of parameters already available would guarantee a valid covariance structure for the enlarged lattice. This is not the case if using the “brute force” approach. If aiming to add another zone, it cannot be guaranteed that the previously estimated parameters would result in a valid covariance structure, hence requiring to re-estimate them. Our various experiments showed that such a new set of parameters would be very close to the previous one. This latter vector of parameters could therefore be used as initial condition of the optimization routine used for their re-estimation.

## 5.2 Choosing an appropriate optimization criterion

When estimating  $\theta$  from data, a discrepancy measure between the imposed GMRF and the observations needs to be chosen. Here estimation is carried out in a maximum likelihood framework. In [40], the authors argue that maximum likelihood estimators for GMRF are not robust with respect to model errors and might result in coefficient estimates which do not describe well the global properties of the data. The authors propose a new optimization criterion which resolves this difficulty. The criterion is based on a norm distance between the estimated and the observed correlation structures.

By considering both the norm-based discrepancy optimization and the maximum likelihood approach, we observed that resulting estimates were very similar. Using

the maximum likelihood approach was then preferred, following another argument in [40], such that, if a GMRF describes the data adequately, then the maximum likelihood-based inference is more efficient.

### 5.3 Parameter estimation using maximum likelihood

Let us focus on a single time  $t$  and recall some of the notations introduced in Sect. 3. An observation of the latent Gaussian field  $\mathbf{x}_t = [x_{t,1}, x_{t,2}, \dots, x_{t,n}]^\top$  is obtained by transformation of the wind power observations  $[y_{t,1}, y_{t,2}, \dots, y_{t,n}]^\top$  through the corresponding predictive cumulative distribution functions, as in Eq. (5). In the case where marginal predictive densities are probabilistically calibrated,  $\mathbf{x}_t$  is a realization from a multivariate Gaussian distribution with zero mean and correlation matrix given by  $\mathbf{Q}^{-1}$ . Consequently the log-likelihood contribution given by  $\mathbf{x}_t$  writes

$$l_t(\theta) = -\frac{1}{2} \left( n \ln(2\pi) - \ln |\mathbf{Q}(\theta)| + \mathbf{x}_t^\top \mathbf{Q}(\theta) \mathbf{x}_t \right). \quad (20)$$

Given  $T$  realizations for the process  $\mathbf{X}_t$ , the overall likelihood is given by

$$l_T(\theta) = \sum_{t=1}^T l_t(\theta) = -\frac{1}{2} \left( nT \ln(2\pi) - T \ln |\mathbf{Q}(\theta)| + \sum_{t=1}^T \mathbf{x}_t^\top \mathbf{Q}(\theta) \mathbf{x}_t \right). \quad (21)$$

Then, solving  $\frac{\partial l_T(\theta)}{\partial \sigma^2} = 0$  with respect to  $\sigma^2$  yields the following profile maximum likelihood estimator for  $\sigma^2$ ,

$$\widehat{\sigma^2} = \frac{\sum_{t=1}^T \mathbf{x}_t^\top \mathcal{P} \mathbf{x}_t}{Tn}, \quad (22)$$

where

$$\mathcal{P} = \begin{pmatrix} \kappa_1 & & & & \\ & 1 & & 0 & \\ & & \rho & & \\ & & & \ddots & \\ 0 & & & & \rho^{K-2} \\ & & & & & \kappa_K \end{pmatrix} \quad (23)$$

Having the profile likelihood estimate for  $\sigma^2$ ,  $\mathbf{Q}$  can be seen as a function of the reduced parameter vector  $\theta^-$ :

$$\theta^- = [\kappa_1, \rho, \kappa_{43}, a_{-1}, b_0, b_{-1}, b_1, c_0, c_{-1}, c_1]^\top. \quad (24)$$

And, finally, an estimate for  $\theta^-$  is obtained by a numerical optimization of the likelihood function given in (21) with respect to the parameter vector  $\theta^-$ .

The requirement for the resulting  $\mathbf{Q}(\hat{\theta})$  to be symmetric positive definite is equivalent to requiring all eigenvalues to be positive. Similarly to [40], we approach the constrained optimization problem as an unconstrained one, adding an infinite penalty if some of the eigenvalues are negative. This approach works well in practice. In parallel,  $\Sigma(\hat{\theta}) = \mathbf{Q}^{-1}(\hat{\theta})$  is required to have a unit diagonal. In practice this is achieved by the corresponding scaling of  $\mathbf{Q}(\hat{\theta})$ , as also suggested in [39].

## 6 Results

### 6.1 Assessing global model fit

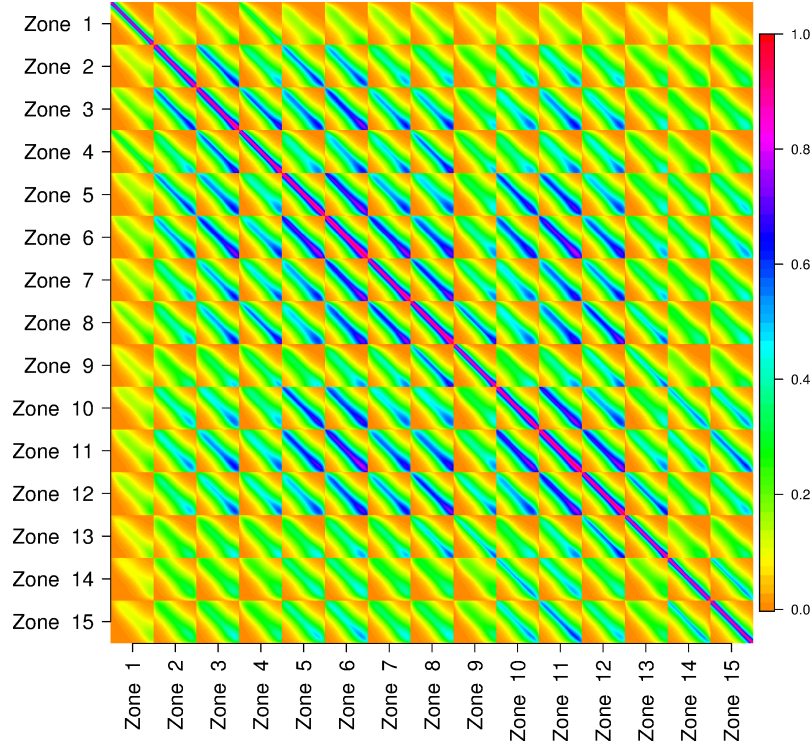
Our verification starts with an examination of the global properties of the estimated dependence structure. This is done in the spirit of [40], i.e., by visually comparing the estimated covariance structure with the sample one, over the first subset of data available and used for modelling and model fitting. The estimated correlation matrix is illustrated in Fig. 8, while the sample one was already shown in Fig. 3. The patterns in these two matrices appear to be very similar.

The motivation for checking the global resemblance between the dependence structures in addition to the overall likelihood evaluation is given by the following. When optimizing the likelihood, the optimal fit is given by fitting the covariances within the neighbourhood exactly, while the remaining ones are determined by the inversion of the fitted precision matrix [40]. This may result in estimates which, instead of capturing dependencies between all the variable pairs in some reasonable way, capture just some of them with a very high precision, while ignoring the others.

### 6.2 Assessing predictive model performance

We then turn ourselves to evaluating the predictive performance of our approach. Indeed, so far, all derivations and discussion were for the first subset of data (between 1st January and 30th November 2006). Here, the evaluation uses genuine forecasts generated for the second subset of data (from 30th November 2006 to 24th October 2007), where the available predictive marginal densities, as well as estimated dependence structure, are used to issue multivariate predictive densities with a dimension  $n = 645$ .

The section starts with a presentation of the benchmark approaches. Further, scores used for the overall quality assessment are discussed. Finally, the empirical results are presented.



**Fig. 8** Estimated correlation matrix

### 6.2.1 Overview of the models considered

The following models are considered in this study:

1. *Independent*: The corresponding multivariate predictive densities are based on the assumption that the marginal densities are independent, i.e.,

$$F_t(y_1, y_2, \dots, y_n) = F_{t,1}(y_1)F_{t,2}(y_2) \cdots F_{t,n}(y_n) ; \quad (25)$$

2. *First order time-dependence*: The corresponding multivariate densities are obtained using a Gaussian copula approach. The covariance matrix accounts only for the temporal dependencies while completely ignoring the spatial ones. This is done by constructing the precision matrix  $\mathbf{Q}$  as described in the above, but setting  $\rho = \kappa_1 = \kappa_K = 1$  and  $b_{-1} = b_0 = b_1 = c_{-1} = c_0 = c_1 = 0$ . That is, the precision matrix in this case is described by the parameters  $a_1$  and  $\sigma^2$  only. This model does not allow for any special arrangement for the boundary points. The conditional precisions are assumed to be constant. In other words, this model corresponds to a conventional stationary GMRF defined by the first order autoregressive process in time;

3. *Separable model with first order decays in time and in space* allowing for non-constant conditional precisions: The corresponding multivariate densities are obtained using a Gaussian copula approach. The precision matrix  $\mathbf{Q}$  is parametrized as in the above while setting  $c_0 = b_0$ ,  $b_1 = b_{-1} = c_1 = c_{-1} = a_1 b_0$ . That is, the precision matrix in this case is described the first order time-dependence (given by  $a_1$ ) and the first order spatial dependence (given by  $b_0$ ). Additionally, the model gives more flexibility compared with the conventional separable covariance structures by considering non-constant conditional precisions (modelled by  $\rho$ ,  $\kappa_1$  and  $\kappa_K$ ). The model does not account for the directional influence, and that is why  $c_j$  is set to be equal to  $b_j$  with all  $j = -1, \dots, 1$ ;
4. *Sample correlation*: The corresponding multivariate predictive densities are obtained using a Gaussian copula approach with the correlation structure given by the sample correlation matrix;
5. *Full model*: The first order model which proposed in this study. That is the precision matrix is described by the full parameter vector  $\theta$  as given in (19).

### 6.2.2 Choosing an appropriate scoring rule for the quality evaluation

In order to evaluate and compare the overall quality of multivariate probabilistic forecasts proper scoring rules are to be employed [8, 17]. An overview of proper scoring rules used for the multivariate forecast verification is given in [18]. In this work the Logarithmic score is used as a lead score for evaluating the performance of the joint predictive densities. The logarithmic scoring rule,  $s$ , is defined as

$$s(p(\mathbf{x}), \mathbf{x}_t) = -\ln(p(\mathbf{x}_t)) \quad (26)$$

Where  $p(\mathbf{x})$  stands for the predictive density, which in our case is given by  $\mathcal{N}(\mathbf{0}, \mathbf{Q}(\hat{\theta})^{-1})$ .  $\mathbf{x}_t$  denotes the corresponding observation.

Suppose, the verification set consists of  $T$  observations, then the overall empirical score value,  $S$ , is given by the average value of the individual  $s(p(\mathbf{x}), \mathbf{x}_t)$ ,

$$S(p(\mathbf{x})) = -\frac{1}{T} \sum_{t=1}^T \ln(p(\mathbf{x}_t)) . \quad (27)$$

In other words, the Logarithmic score value over the evaluation set is given by the average minus log-likelihood derived from the observations. The score is negatively orientated: the lower its value is, the better.

There are several reasons for choosing the Logarithmic score as the lead evaluation criterion. Firstly, it is consistent with the optimization criterion used when estimating the model parameters. Secondly, allowing for some affine transformations, this is the only local proper score (see Theorem 2 in [2]). Locality means that the score depends on the predictive distribution only through the value which the predictive density attains at the observation [8]. An important advantage of using local scores when dealing with multivariate predictive densities comes with the re-



lated computational benefits. When dealing with local scores, there is no need to draw random samples from the predictive density in order to make the evaluation.

For instance, an alternative is to use the Energy score (see detailed information on this in [18]). This score is non-local and is based on the expected Euclidean distance between forecasts and the related observations. Most often, closed-form expressions for such expectation are not available. One then needs to employ Monte-Carlo methods in order to estimate the score value [18]. In high dimensions, Monte-Carlo techniques result in computational challenges.

A downside of employing local scores is their sensitivity to outliers. For instance, the Logarithmic score is infinite if the forecast assigns a vanishing probability to the event which occurs. In practice, when working with real data, such sensitivity might be a problem.

In this work, we considered both the Energy score and the Logarithmic score for the final density evaluation. In general the results suggested by the two scores were consistent and no contradictions were observed. However, we noticed that the Energy score was not very sensitive to the changes in the correlation structure. That is, the changes in the Energy score when moving from the assumption of independence between the marginal predictive densities to models accounting for the dependence structure were rather small (even though they still proved statistically significant based on Diebold-Mariano test statistics [12]). This is caused by low sensitivity of the Energy score to changes in the dependence structure as argued in [38]. This is another reason to focus on the Logarithmic score further in this study.

### 6.2.3 Empirical results

The results from our evaluation of multivariate predictive densities issued based on the approaches mentioned in Sect. 6.2.1 are collated in Table 1, while also describing the complexity of these models in terms of their number of parameters.

**Table 1** Quality assessment of the predictive densities in terms of the Logarithmic score ( $S$ )

Model	Nr. of parameters	S
Independent	0	853.14
First order in time	1	409.98
Separable space-time model	6	357.84
Full model	10	318.07
Sample correlation	207690	267.96

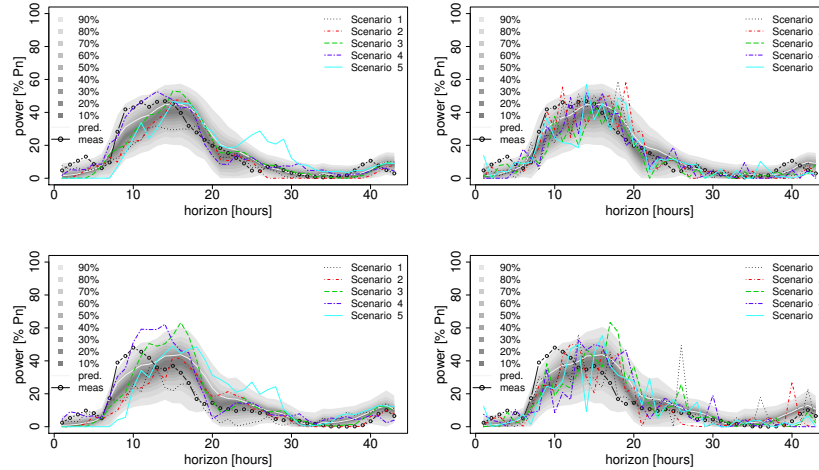
One can appreciate the importance of accounting for the dependence structure from the fact that multivariate predictive densities derived from the independence assumption are shown to be of lowest quality. The full model proposed in this study outperforms the other two dependence structures, i.e., both the first-order time-dependence and the separable space-time model. The statistical significance of the

improvements was verified using a likelihood-ratio test, similar to that described in [27]. This confirms that letting the related conditional correlations change depending on the direction as well as allowing for non-separable space-time influence results in better quality of the multivariate probabilistic forecasts.

Predictive densities defined by the sample correlation matrix provide the best quality forecasts. This is also expected, since in this study the estimation period consisted of one year of hourly data. Large amount of data made it possible to estimate the covariance structure, even for such a high dimension. However, the main interest in the future is to make the covariance structure dependent on meteorological conditions. In this setup, tracking a sample covariance will become nearly impossible. Thus, the proposed parametrization is crucial for further development of the methodology as it significantly reduces the effective problem dimension.

### 6.3 Scenario generation

As an illustration of probabilistic forecasts obtained with the proposed approach, Fig. 9 depicts a set of five scenarios describing wind power generation at zones 6 and 7, for lead times between 1 and 43 hours ahead, issued on the 15th of June 2007 at 01:00. The marginal predictive densities originally available are also shown, as well as the power observations obtained a posteriori.



**Fig. 9** Scenarios describing wind power generation at zones 6 (top) and 7 (bottom) from 1 to 43 hours ahead issued on the 15th of June, 2007, at 01:00. The scenarios shown in the left are those obtained with our model. Those on the right are obtained under an independence assumption, thus, not respecting neither temporal nor spatial dependencies

The scenarios generated using our approach respect dependencies both in time and in space. Respecting correlations in time ensures that the corresponding scenarios do not evolve as a random walk whose magnitude would be shaped by predictive marginals. For instance, given that a scenario predicts wind power generation at time  $t + k$  to be far from its conditional expectation, then power generation at time  $t + k + 1$  is also expected to deviate substantially from its conditional expectation. As an example see Scenario 5 for zone 6 for lead times between 22 to 30 hours ahead. Similarly, respecting spatial dependency between the zones ensures that when large (resp. small) forecast errors are observed at one zone, the errors at the other zone are also expected to be large (resp. small). This is also visible from Fig. 9. For example, in the case of scenario 4, wind power generation deviates a lot from its conditional expectation in both zones 6 and 7, for lead times between 22 to 30 hours ahead. In contrast, the corresponding scenarios generated using the independent model do not respect neither temporal, nor spatial dependencies in the data.

## 7 Concluding remarks and perspectives

The problem of obtaining multivariate predictive densities of wind power generation was considered, jointly for a number of contiguous spatial areas and lead times, based on already available marginal predictive densities for each individual zone and lead time. A Gaussian copula approach was employed to build such multivariate predictive densities. Our core contribution lies in the proposed parametrization of the dependence structure. More specifically, instead of modelling the covariance matrix directly, focus is given to its inverse (precision matrix). This solution brings substantial practical benefits. For one, the precision matrix is shown to be very sparse. This permits to place ourselves in GMRF framework, hence resulting in computational benefits obtained through faster factorization algorithms available for sparse matrices. Besides, the proposed parametrization allows for more flexibility, since one may readily obtain nonseparable covariance structures.

The data analysis carried out for the Danish dataset revealed that the empirical precision matrix shows non-constant conditional precisions (increasing with the lead time), as well as varying conditional correlations. This hence required to go beyond conventional approaches relying on homogeneous stationary Gaussian fields. We proposed a way to model changes in conditional precisions, also allowing for conditional correlations to change with the direction. Accounting for such directional influence is not only necessary when looking at the data, but it is also quite intuitive, provided that wind power forecast errors propagate in time and in space under the influence of meteorological conditions. Consequently, the application results in terms of predictive performance confirmed that the proposed methodology and precision matrix parametrization could yield benefits in generating high-dimensional multivariate predictive densities of wind power generation, illustrated by lower Logarithmic score values.

Besides the methodological proposal and application results, a number of relevant challenges and perspectives for future work were identified. Firstly, a direct extension of the proposed methodology could consist in conditioning the precision matrix on meteorological conditions. Here, for simplicity, it was considered that the dependence structure was constant through time, with a stationarity assumption for the underlying process. In practice, however, such a spatio-temporal dependence structure may vary substantially. There are many factors which might influence changes in process dynamics. An obvious one is the influence of the surface and higher-level wind conditions. The influence of wind direction on the spatio-temporal dependence structure could be readily modeled with a regime-switching approach, by allowing the neighbourhood structure to change with wind direction. In other words, instead of distinguishing between “West-East” and “North-South” neighbourhoods, one could instead consider “Upwind”-“Downwind” and “Concurrent”-“Concurrent”. Also, it would be interesting to investigate ways to explain the variations in the conditional precisions among the zones. Possibly some clustering techniques could be employed. In parallel, slow variations in the process dynamics could be captured by considering adaptive estimation schemes for the precision matrices.

Further, an interesting challenge will be to move from the lattice setup considered in this study to a fully continuous approach. Following [25], the link between stochastic partial differential equations and some type of precision matrices could be used for such a type of problem. By understanding how the elements of the precision matrix evolve with distance between the zones and prevailing meteorological conditions, one can get a process description via stochastic partial differential equations.

On the forecast verification side, a clear challenge relates to the high dimension of the multivariate predictive densities. Already when working with a dimension  $n = 645$ , we have faced both methodological and computational issues, in view of the different scoring rules available for multivariate probabilistic forecast verification. Even though both Logarithmic and Energy scores can be used for multivariate probabilistic forecast verification, each of them introduces limitations in the verification exercise. On the one hand, the Energy score, being a non-local score, comes with additional computational costs since its estimation requires Monte Carlo techniques. Furthermore, following [38], this score has low sensitivity to changes in covariance structure. On the other hand, the Logarithmic score is highly sensitive to outliers: this may clearly cause difficulties in practical applications, for which both noisy data and model misspecification may then become problematic. Overall, the field of multivariate probabilistic forecast verification needs increased focus in order to propose theoretically sound and practical ways to thoroughly evaluate high-dimensional forecasts such space-time trajectories of wind power generation.

## Acknowledgements

The authors were partly supported by the EU Commission through the project SafeWind (ENK7-CT2008-213740), which is hereby acknowledged. Pierre Pinson was additionally supported by the Danish Strategic Research Council under ‘5s’–Future Electricity Markets (12-132636/DSF). The authors are grateful to Energinet.dk, the transmission system operator in Denmark, for providing the observed power data used in this paper, and to ENFOR A/S for generating the point forecasts of wind power generation used as input. Finally, reviewers and editors are acknowledged for their comments and suggestions on an earlier version of the manuscript.

## References

1. T. Ackermann et al. *Wind power in power systems*. John Wiley & Sons: New York, 2005.
2. J. M. Bernardo. Expected information as expected utility. *Ann. Stat.*, 7(3):686–690, 1979.
3. J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc. B*, pages 192–236, 1974.
4. R. J. Bessa, M. A. Matos, I. C. Costa, L. Bremermann, I. G. Franchin, R. Pestana, N. Machado, H. P. Waldl, and C. Wichmann. Reserve setting and steady-state security assessment using wind power uncertainty forecast: a case study. *IEEE Trans. Sustain. Energy*, 2012.
5. R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang. Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renew. Energ.*, 40(1):29–39, 2012.
6. A. Botterud, Z. Zhou, J. Wang, R. J. Bessa, H. Keko, J. Sumaili, and V. Miranda. Wind power trading under uncertainty in LMP markets. *IEEE Trans. Power Syst.*, 27(2):894–903, 2012.
7. A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J. Bessa, and V. Miranda. Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of illinois. *IEEE Trans. Sustain. Energy*, 4(1):250–261, 2013.
8. J. Bröcker and L. A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecast.*, 22(2):382–388, 2007.
9. E. D. Castronuovo, I. Sánchez, J. Usaola, R. Bessa, M. Matos, I.C. Costa, L. Bremermann, J. Lugaro, and G. Kariniotakis. An integrated approach for optimal coordination of wind power and hydro pumping storage. *Wind Energy*, 2013. available online.
10. A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renew. Sust. Energ. Rev.*, 12(6):1725–1744, 2008.
11. G. Diaz. A note on the multivariate archimedean dependence structure in small wind generation sites. *Wind Energy*, 2013. available online (DOI: 10.1002/we.1633).
12. F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *J. Bus. Econ. Stat.*, 13(3):253–263, 1995.
13. C. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, 12(4):347–368, 2007.
14. G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power—A literature overview. 2nd edition. Technical report, Technical University of Denmark, 2011.
15. R. Girard and D. Allard. Spatio-temporal propagation of wind power prediction errors. *Wind Energy*, 2012.
16. T. Gneiting. Editorial: probabilistic forecasting. *J. R. Stat. Soc. A*, 171(2):319–321, 2008.
17. T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.*, 102(477):359–378, 2007.

18. T. Gneiting, L. I. Stanberry, E. P. Gritmit, L. Held, and N. A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, 17(2):211–235, 2008.
19. C. M. Hafner and H. Manner. Conditional prediction intervals of wind power generation. *J. Appl. Econom.*, 27(2):269–295, 2012.
20. S. Hagspiel, A. Papaemannouil, M. Schmid, and G. Andersson. Copula-based modeling of stochastic wind power in europe and implications for the swiss power grid. *Appl. Energ.*, 96:33–44, 2012.
21. M. Hofert, M. Mächler, and A. J. Mcneil. Likelihood inference for archimedean copulas in high dimensions under known margins. *J. Multivar. Anal.*, 110:133–150, 2012.
22. J. Jeon and J.W. Taylor. Using conditional kernel density estimation for wind power density forecasting. *J. Amer. Statist. Assoc.*, 107(497):66–79, 2012.
23. L. Jones and C. Clark. Wind integration - A survey of global views of grid operators. In *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.
24. M. Kyung and S. K. Ghosh. Maximum likelihood estimation for directional conditionally autoregressive models. *J. Stat. Plan. Infer.*, 140(11):3160–3179, 2010.
25. F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. B*, 73(4):423–498, 2011.
26. H. Louie. Evaluation of bivariate archimedean and elliptical copulas to model wind power dependency structures. *Wind Energy*, 17(2):225–240, 2014.
27. H. Madsen and P. Thyregod. *Introduction to general and generalized linear models*. Chapman & Hall/CRC Press, 2011.
28. A. Möller, A. Lenkoski, and T. L. Thorarinsdottir. Multivariate probabilistic forecasting using bayesian model averaging and copulas. *Q. J. R. Meteorol. Soc.*, 139:982–991, 2013.
29. H.Aa. Nielsen, T.S. Nielsen, and H. Madsen. An overview of wind power forecasts types and their use in large-scale integration of wind power. In *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.
30. M. A. Ortega-Vazquez and D. S. Kirschen. Assessing the impact of wind power generation on operating costs. *IEEE Trans. Smart Grid*, 1(3):295–301, 2010.
31. G. Papaefthymiou and D. Kurowicka. Using copulas for modeling stochastic dependence in power system uncertainty analysis. *IEEE Trans. Power Syst.*, 24(1):40–49, 2009.
32. A. Papavasiliou and S. S. Oren. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. *Oper. Res.*, 61:578–592, 2013.
33. P. Pinson. Wind energy: Forecasting challenges for its optimal management. *Stat. Sci.*, 28(4):564–585, 2013.
34. P. Pinson and R. Girard. Evaluating the quality of scenarios of short-term wind power generation. *Applied Energy*, 96:12–20, 2012.
35. P. Pinson and G. N. Kariniotakis. Conditional prediction intervals of wind power generation. *IEEE Trans. Power Syst.*, 25(4):1845–1856, 2010.
36. P. Pinson, H. Madsen, H. Aa. Nielsen, G. Papaefthymiou, and B. Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.
37. P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10(6):497–516, 2007.
38. P. Pinson and J. Tastu. Discrimination ability of the Energy score. Technical report, Technical University of Denmark, 2013.
39. H. Rue and L. Held. *Gaussian Markov random fields: theory and applications*, volume 104. Chapman & Hall, 2005.
40. H. Rue and H. Tjelmeland. Fitting gaussian markov random fields to gaussian fields. *Scand. J. Stat.*, 29(1):31–49, 2002.

41. D. Simpson, F. Lindgren, and H. Rue. In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics*, 23(1):65–74, 2012.
42. M. Sklar. *Fonctions de répartition à  $n$  dimensions et leurs marges*. Université Paris 8, 1959.
43. J. Tastu, P. Pinson, and H. Madsen. Multivariate conditional parametric models for a spatiotemporal analysis of short-term wind power forecast errors. In *Scientific Proceedings of the European Wind Energy Conference, Warsaw*, pages 77–81, 2010.